THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 62, No. 7, September 1983
Printed in U.S.A.

Total Network Data System:

Small Office Network Data System

By D. H. BARNES* and J. J. O'CONNOR*

(Manuscript received December 3, 1982)

The Small Office Network Data System (SONDS) is a part of the Total Network Data System (TNDS) designed specifically to meet the network data requirements for the small step-by-step offices of the Bell System. Its ability to accurately relate to the load and service characteriatics seen in these small, mostly rural officea has ensured its success. It maintains low cost by meaauring and recording a aingle observation per day for each traffic measurement. Tbis is the daily peak hour load per group. A moving average of the mean and variance of peak hour measurements is interpreted by using a modified-Gumbel extreme value distribution function. This has been demonstrated to fit well within the actual distributions of daily peak hour loads. It is used to construct traffic capacity tables that better reflect customers service experience in these offices than do existing methods. It uses field-tested parameter assumptions of the underlying distribution from which peak valuea are sampled to estimate equivalent time-consistent busy hour loads relative to the observed peak hour loads. This allows service to be compared among SONDS and non-SONDS step-by-step offices. SONDS uses on-site data collection units, a single central computer for all Bell System offices, and provides local printing of output reports. Data are collected each night and daily, weekly, and monthly output reports are distributed over the switched network.

* Bell Laboratoriea.

I. INTRODUCTION

The Small Office Network Data System (SONDS) is a miniature Total Network Data System (TNDS) designed to meet the unique and special needs of the Bell System's smaller, rural-type offices. In respect to SONDS a small office is one serving fewer than 5000 main stations (MSs). These comprise almost 40 percent of the Bell System's total switching entities but serve less than 8 percent of its customers. The offices that are candidates for SONDS are predominantly of the noncommon control, step-by-step type. They are generally located in remote rural locations and are unattended.

The amount of traffic and the service vary among these offices and among the serving groups within each office for several reasons. First, as random traffic theory predicts, there is a greater proportional variation in traffic with small versus large groupings of traffic sources. Observations have shown that the actual variations of source traffic are, in fact, much greater than assumed by the random theory underlying the probability tables normally used to size central offices. Second, the mix of customer types varies widely among these offices and this further affects the within-a-day and day-to-day traffic distributions in often unpredictable ways. In addition, as a result of the relatively small originating line and terminating connector groups of the small step-by-step offices, load balancing problems are accentuated. These offices are unattended, which means less frequent surveillance of central office equipment problems such as busy switches or trunks. On the average, there will be more such maintenance problems and they will be longer in duration than in larger, attended offices.

The small offices require special attention if both service and cost objectives are to be met. They are, however, too small and remote to justify support by conventional, continuous measurement systems. Typically, a few days of traffic data per year are available in these offices. Present standard dimensioning techniques and design service criteria are thus inadequate in the presence of minimal data and large traffic variation. We have demonstrated with SONDS how large office standards can be applied to these small offices at an affordable price.

II. DIMENSIONING OF TRAFFIC SYSTEMS

Design service criteria for standard Bell System engineering are based on the concept of a Time-Consistent Busy Hour (TCBH). It is by definition that clock hour which has been observed by study to contribute the greatest amount of total traffic over some defined period of time. The sizing of a group of traffic servers is based on meeting service criteria for the highest controlling TCBH loads for the average busy season, average ten high day, or highest day.

The traffic theory used to calculate engineering capacity tables assumes the behavior of a constant system of chance causes ("statistical equilibrium"). To account for different levels of departure from truly random behavior, traffic capacity tables are adjusted to allow for options of low, medium, or high levels of day-to-day variation.

Time-consistent busy-hour practice is a compromise for data collection and processing costs. Simplifying assumptions are used, however, which lead, particularly in small offices, to overprovisioning to protect for variations not reflected in an average statistic. Due to lack of data, maintenance problems, and forecasting errors, underprovisioning can also occur.

TCBH practices assume that provision of objective service in the average busy hour will provide adequate service in all hours. Average service is not, however, the service seen by individual customers nor by groups of customers in offices with traffic imbalance. While TCBH service criteria have been sbown by experience to be a reasonable approach to provide an adequate overall service control for large groups of customers, they are less valid for small groups because of tbe greater deviation from truly random behavior of such groups. For example, SONDS office studies reveal that the TCBH is the actual busiest hour of the day on only about 20 percent of the days.

III. THE SONDS DEVELOPMENT

SONDS is the outcome of a series of Bell Laboratories' traffic studies and computer experiments that were initiated originally to explore the power of computers to completely control data collection from a central location. Included were studies of traffic characteristics and the evaluation of design service criteria that might be used to better control service levels and at the same time provide a more efficient allocation of resources. In recognition of the largely unsolved small office data and design problem, we initiated an experimental network data system designed to meet that need. These relatively simple offices, largely unmeasured, were judged to be a good environment to explore a completely mechanized system that would use time-shared computers to interface with smart microprocessor-controlled central office terminals. We were looking for a system free from the control of people other than the direct user. These offices were not directly coupled to TNDS and thus gave us a simpler environment in which to study innovative traffic systems.

The SONDS development, the formulation of analysis requirements, and the specifications of output reports were designed to incorporate:

- Low cost
- Service measurement
- A compact database

- A nearly paperless system
- Switched-network data communication
- An interactive computer interface
- Automatic report generation and distribution
- Complete access by the authorized user for:
  —Nonscheduled data or special study requests
  —Database control
  —Trouble reports and data for analysis
  —Remote testing of terminal equipment.

SONDS has met these desired objectives. Its success is due to a combination of the following features:

1. An Extreme Value Engineering (EVE) method has been formulated and adapted to the studied characteristics of the candidate offices. EVE uses order statistics of an extreme value distribution formulated from daily peak hour data observations. The need for simplifying assumptions using otber traffic dimensioning models is largely reduced. Through use of tbe order statistics the computer is able to detect questionable data and to issue timely exception reports, including backup data, which are useful for analyzing and correcting trouble conditions.

2. New EVE design service criteria are based on a broad study of office characteristics and are selected to match the objectives of current dimensioning methods. On the average, with the same or less switching equipment and trunks, overall service will improve. With full-time data surveillance available for the first time and with a good system of trouble detection, there will be less need to play safe by overprovisioning.

3. A compact database is realized because only a single hourly value per day, the daily peak hour value, is recorded in the database for each traffic measurement. The extreme value distribution parameters are estimated from a moving average of the mean and variance of this measurement.

4. Automatic computer polling of office terminals during early morning hours is an efficient serial use of computer ports and the switched network.

5. A largely paperless system has been realized. The standard reports required in TNDS are automatically distributed immediately following the end of each reporting interval. Database updates and requests for special data and reports are made via an interactive interface with the computer and require no other (human) interactions.

6. No special monitoring or scaling is required to compensate for holiday or "odd-ball" traffic conditions in the office engineering because automatic outlier tests reject such data.

7. Dial tone delay service measurements and their reporting as a service index are contained in the system. Separate and costly dial tone test call facilities are not needed. The result is a more accurate measurement and includes a measure of group-to-group service differences not available otherwise.

IV. FORMULATION OF THE EVE METHOD FOR SONDS OFFICES

The following is a short history of work that helped generate an interest in the EVE approach used in SONDS.[1–4]

4.1 Early work

Karlsson[1] claimed that, particularly for international routes, the (time-consistent) busy hour is both unstable in amplitude and unfixed in time. He proposed the dimensioning of traffic routes from integrated information of all traffic peaks taken over a sufficiently long time interval T, typically six hours a day for five days, to give good statistical stability. This was an early attempt to recognize the importance of differences in the variances of traffic groupings with equal loads.

Gershwin, Laue, and Wolman[2,3] solved a problem related to the load administration of Subscriber Line Concentrators (SLC) using extreme value statistics. The problem was to administer line assignment far enough ahead so that concentrated lines could see nearly zero blocking in this additional switching stage. Extreme value methods were used to keep peak hour blocking less than 0.005 except for a few times per year.[4]

Unlike Karlsson, Gershwin et al. retained hourly measurements but dropped the concept of a time-consistent busy hour. They recorded only a weekly peak load. The mean and variance of these weekly peak loads were found to be good estimates of the fitting parameters of the Gumbel extreme value distribution function.[5] The plan was simple and inexpensive and served well to direct line assignments and to provide capacity estimates for planning purposes.

4.2 Daily peak values

The initial work that led to extreme value engineering as used by SONDS is published in Refs. 6, 7, and 8. In SONDS it is important that traffic reports and service measurements be integrated into and be consistent with the rest of TNDS. It was thus important, in contrast to weekly values used for the SLC, to follow daily busy hour and busy season concepts and to furnish monthly traffic reports consistent with other TNDS reports. The use of daily peak values for design and administration of small offices was of particular interest because we had experimented with techniques to measure, collect, and summarize such data at a very low cost.

Day-to-day variation of TCBH originating traffic loads as observed in step-by-step offices is illustrated in Fig. 1. The coefficient of variation in percent is shown as a function of offered load in erlangs. Curve A is an approximate fit to the TCBH loads observed in many step-by-step offices. The data showed considerable scatter, indicating that day-to-day variation is more than just a function of the offered traffic. Curve C is a theoretical curve based on the assumption of constant system of chance causes ("statistical equilibrium") whicb is basic to the probability theory used to construct traffic capacity tables. Curve B depicts a possible location for daily peak-bour data. Its location between the theoretical curve C and the TCBH curve A will be a function of the number of hours in the day that have a load high enough to possibly be the peak load for tbe day. For calculation purposes we will develop an equivalent number of equally loaded hours with which to describe the underlying distribution from which peak hours come. We designate these as "Candidate Busy Hours".

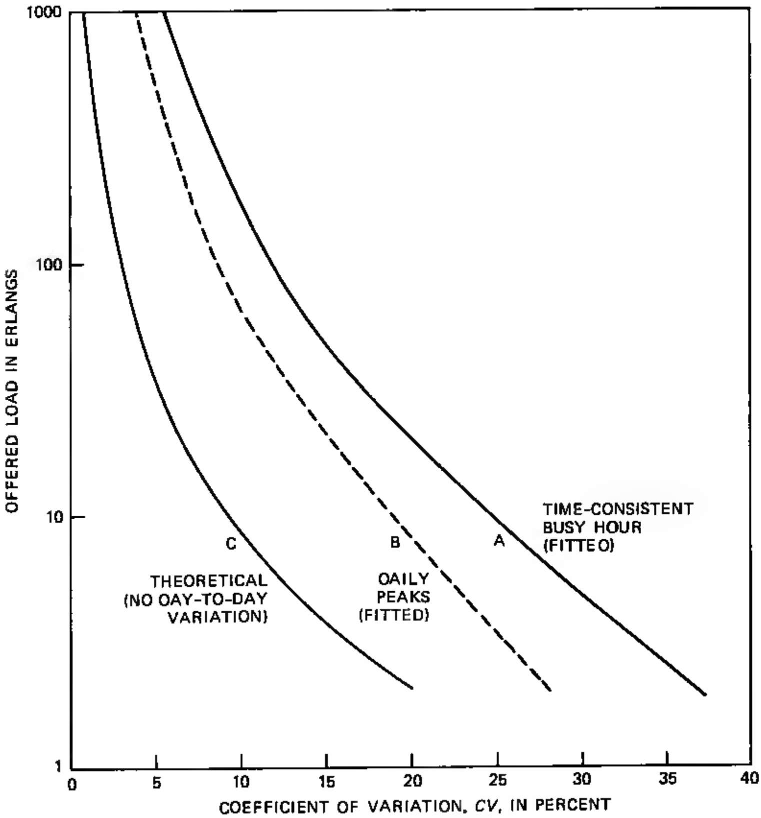

Fig. 1—Variability of hourly loads.

Figure 2 illustrates a comparison between TCBH and peak-hour concepts using actual data. These data were taken from a typical line finder group in a SONDS office, which had 13 line finders serving the group. The data are plotted on probability paper constructed to match the normal-to-the-sixth extreme value probability function derived in Section 4.4.

Curve A is a theoretical curve based on purely random, no day-to-day variation; it is the underlying distribution of the TCBH capacity tables. We could not plot this curve exactly as a straight line on this graph. To simplify the illustration we have approximated it as two line segments plotted between the 0.05 and the 0.95 probabilities points and the average value. This average value was scaled to be equal to the 242 hundred call seconds (CCS) measured average of the TCBH data. This gives the 0.95 point, equivalent to a 20-day return period load, a load of 286 CCS.

Curve B is a plot of the 20 TCBH loads observed in this office. It is seen that these data points significantly deviate from the theoretical curve A, and this is due in part to actual day-to-day variations. Note that the observed 20-day return period load, the 0.95 point, is 326 CCS during the time-consistent busy hour.

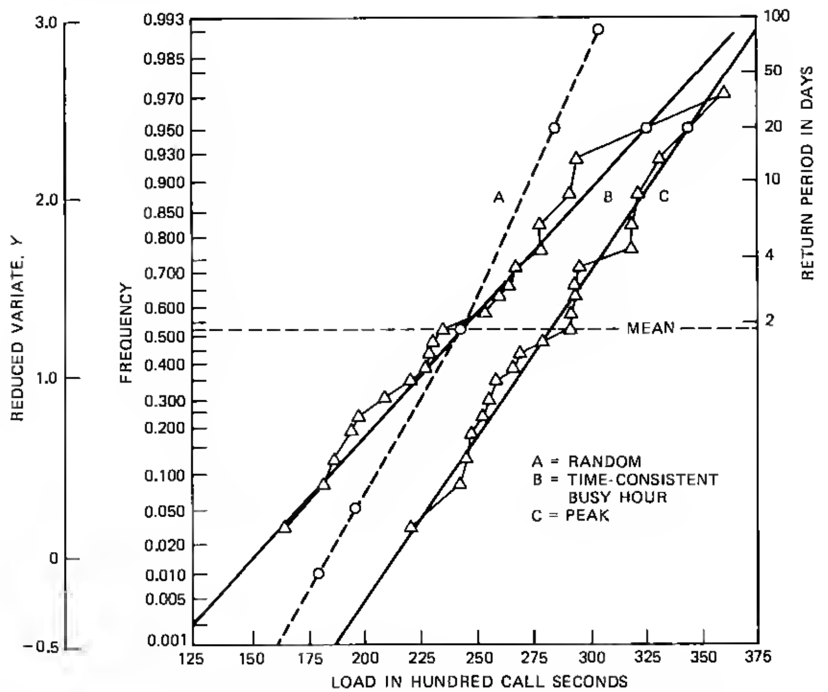

Fig. 2—Extreme value probability plot for originating loads.

Curve C is a plot of the 20 daily peaks taken for the same days as the TCBH data. The straight-line character of the data on this graph is typical of the agreement between daily peak data and the normal-to-the-sixth distribution function. Note that the mean value is 283 CCS and the 0.95 value is 346 CCS. While the mean and 0.95 values of the peak curve are larger than those of the TCBH curve, the larger slope of the peak curve indicates less variance, that is, daily peak values are less volatile than TCBH values!

These several load values obtained from Fig. 2 can be used to compare the dial tone delay service as given by current engineering capacity tables with the range of expected values that a customer will experience. This is given in Table I, which shows the service values for both mean and 20-day return period loads. Since the 20-day return period load X_{20} is also interpreted as the load to be exceeded once a month, on the average, we can examine the service implications on this basis. Table I shows that a line finder group properly designed to give 1.5-percent dial tone delay during the time-consistent busy hour would be expected to exceed 4.3 percent once a month, even with purely random traffic. The day-to-day variation observed in Fig. 2 would increase this to 8.7 percent, and the use of the daily peak-hourly loads of Fig. 2 would predict a service level of 12.5 percent to be exceeded once a month.

Extreme value statistical methods can be applied to load balance. Loads are balanced through line assignment, which takes into account both additions and disconnects. Table II shows an office TCBH and peak load averages and their coefficients of variation, *CV*, using a month of data. The last column, X_{20}, is the 20-day return period load calculated from the mean and standard deviation of the daily peak values. As expected, we can see that the *CV* of the daily peak loads (12 percent) is less than that of the TCBH loads (22 percent). Also note that the *CV* of the peak loads (6 percent) across the loading division is less than that of the TCBH loads (7.3 percent). Even the X_{20}, once-a-month load, which includes the additional parameter of variance, has only a *CV* of 7.9 percent.

Table III shows how the corrective action needed to balance load against load main station movement depends on the type of data used

Table I—Load service characteristics

	Load in CCS		Expected Dial Tone Delay (13 Line Finders)	
Distribution	Mean	X_{20}	Mean	X_{20}
A–Random	242	293	1.5	4.3
B–TCBH	242	326	2.1	8.7
C–Peak BH	283	346	4.2	12.5

Table II—Originating load distributions

Line Finder Group No.	TCBH CCS	TCBH CV (Percent)	Daily Peaks CCS	Daily Peaks CV (Percent)	X_{20} CCS
1	159	21	185	11	220
2	155	19	199	8	226
3	139	17	185	8	211
4	179	17	214	9	249
5	169	25	220	14	273
6	147	21	200	13	244
7	162	22	196	12	238
8	164	26	193	18	252
9	162	29	207	13	253
Avg	160	22	200	12	241
Total percent of CV		7.3		6.0	7.9

Table III—Originating main station balance

Line Finder Group No.	TCBH	Daily Peaks	X_{20}
1	−1	−15	−18
2	−6	−1	−13
3	−27	−15	−25
4	+24	+14	+7
5	+12	+20	+27
6	−17	0	+3
7	+3	−4	−3
8	+5	−7	+9
9	+3	+7	+10
Avg.	11	9	12

from Table II. The best service balance will be accomplished using X_{20} data. Note the very large differences by individual groups between the movement indicated by TCBH and that indicated by X_{20} data.

4.3 Gumbel distribution

Preliminary work on SONDS, following the direction of the previous work, used the Gumbel[5] extreme value distribution, a double exponential distribution, shown below.

$$G(x) = \exp\{-\exp[-\alpha(x - \mu)]\}. \tag{1}$$

$G(x)$ is the probability that the largest observation in a large set of observations is less than the value x. In this expression μ is a location parameter (the mode), and α is an inverse measure of dispersion. The mean of the distribution is:

$$E(x) = \mu + \gamma/\alpha, \tag{2}$$

where $\gamma = 0.57721$ (Euler's Constant) and the variance is:

$$V(x) = \frac{\Pi^2}{6\alpha^2}. \tag{3}$$

The Moments Method may be used to estimate μ and α by replacing $E(x)$ and $V(x)$ with their sample values from a sample of N hourly peaks $x_1, \cdots, x_N$. The sample mean is:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{4}$$

and the sample variance is:

$$\bar{s}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2. \tag{5}$$

Substituting eqs. (4) and (5) into eqs. (2) and (3), we have:

$$\hat{\mu} = \bar{x} - \gamma/\hat{\alpha} \tag{6}$$

and

$$\hat{\alpha} = \frac{\Pi}{\bar{s}\sqrt{6}}. \tag{7}$$

Data points fitting a Gumbel distribution function should plot as a straight line on Gumbel extreme value probability paper. Figure 3 shows an attempt to do this with hourly peak data obtained in a Private Branch Exchange (PBX) office serving a single business customer.[6] Notice that the data points do not fall within the control limits. A number of the low peaks were identified as holidays and days preceding holidays. By applying rejection test statistics to one low point at a time and recalculating the order statistics until the lowest point fell within control limits, it was possible to plot the data as shown in Fig. 4, after rejecting only nine of the original 80 data points. This method appeared to properly represent the high data points of most interest and automatically reject data of little interest, such as holidays.

If we reject the lowest few points to force a straight line fit of data to the double exponential probability plot, which is useful for eliminating holiday traffic, we also might eliminate some low data points necessary to identify problems in SONDS. Tbis forced rejection of the lower data points makes the estimate of the mean of the daily peaks biased on the high side and the estimate of the variance biased on the low side, making the effect on the resulting estimated 20-day return period load indeterminate.

We realized that the problem arises because the daily peak loads do not satisfy the large sample assumption that leads to the double

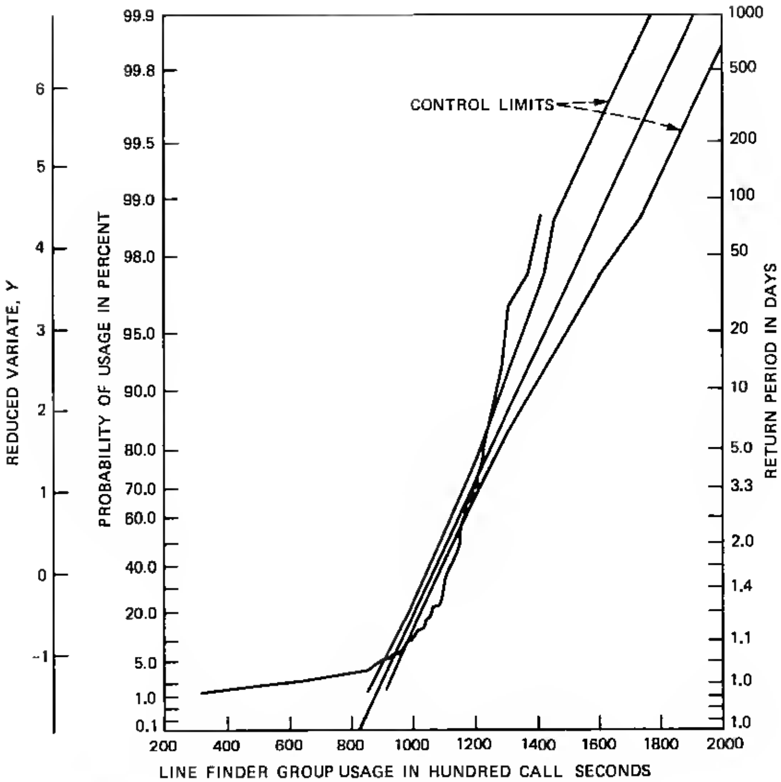

Fig. 3—Extreme value probability plot for 80 peak hours of data.

exponential distribution function of eq. (1). Instead of selecting a peak value from among a large number of candidate hours, we find that we are typically selecting daily peaks from a few to perhaps ten candidate busy hours.

4.4 Normal-to-the-sixth distribution

An improved extreme value distribution function model was needed to fit the small office traffic data. A set of statistical test procedures was needed in the initial start-up testing of a SONDS office (see Section 9.1) to resolve installation problems and other central office problems not previously detected for lack of data.

The first step was to test the distribution of equipment loads across days for the busier hours of each of many offices. The findings were that the distributions satisfactorily tested for normality. The second step was to test whether round-the-clock hourly data can be modeled adequately by h equally loaded candidate busy hours even though the actual clock hours do not have equal loads. These tests led to the

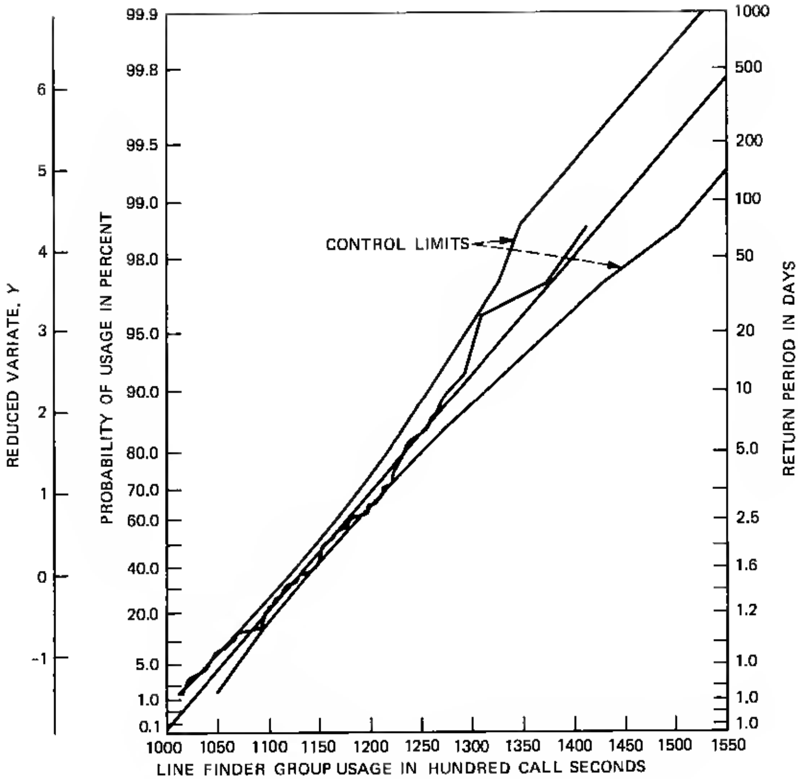

Fig. 4—Extreme value probability plot for 71 peak hours of data.

conclusion that a modified extreme value distribution function suitable for SONDS data is the normal probability distribution function raised to an optimum power of h. Table IV gives the results of three tests that resolved the optimum value of h as 6.

Therefore, the modified extreme value distribution function for describing daily peak-hour usage data, x, is the normal distribution raised to the sixth power, $F(x)$:

$$F(x) = \left[\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x} \exp[-(t-\mu)^2/2\sigma^2]dt\right]^6, \tag{8}$$

where μ and σ are the mean and standard deviation of the underlying equal "candidate busy hours" load distributions.

Tippett[9] numerically evaluated the probability of finding the largest value of h samples drawn from a normal distribution. In particular for $h = 6$, in terms of the normalized variate, the mean is:

$$E(x) = \mu + 1.267\sigma \tag{9}$$

Table IV—Selection of number of candidate busy hours, h, in the (NORMAL)^h model

Tests	Ranges of h Indicated by Tests
Distribution of daily peak loads among hours during the day	6–10
Validation and monitoring of lower daily peak loads	4–6
Estimating highest of 20 daily peaks	6–9
Collectively (all of the above)	4–10
Value selected	6

and the variance is:

$$\mathrm{Var}(x) = 0.416\sigma^2. \tag{10}$$

In SONDS using the moments method, μ and σ are estimated by matching the first two moments of the distribution of eq. (8) to their unbiased sample estimates. Replacing $E(x)$ and $\mathrm{Var}(x)$ by the unbiased sample estimates of the mean, $\bar{x}$, and variance, $\bar{s}^2$, the moments method equations are obtained:

$$\bar{x} = \hat{\mu} + 1.267\hat{\sigma} \tag{11}$$

and

$$\bar{s}^2 = 0.416\hat{\sigma}^2. \tag{12}$$

Manipulating eqs. (11) and (12), we have the parameters of the normal-to-the-sixth-distribution in terms of the underlying daily peak parameters as

$$\hat{\mu} = \bar{x} - 1.96\bar{s} \tag{13}$$

and

$$\hat{\sigma} = 1.55\bar{s}, \tag{14}$$

where $\bar{x}$ and $\bar{s}$ are the unbiased sample estimates, calculated directly from the daily hourly peaks x_i, as shown in eqs. (4) and (5).

The 95th percentile of $F(x)$ can be calculated directly from eq. (8) by setting $F(x) = 0.95$ to give

$$x_{20} = \hat{\mu} + 2.39\hat{\sigma}, \tag{15}$$

which by eqs. (13) and (14) can be stated in terms of the daily peak values as

$$x_{20} = \bar{x} + 1.74\bar{s}. \tag{16}$$

This x_{20} is the load that is exceeded once, on the average, every 20 business days. It is also called the once-a-month load.

V. EVE SERVICE CRITERIA

The use of daily peak-hour measurements to estimate a once-a-month load reflects to a greater degree the service viewpoint of customers than do present average time-consistent busy hour practices. The estimated once-a-month load is based on continuous monitoring of all hourly data, and as interpreted in SONDS by the $F(x)$ model of eq. (8), takes into account the variability of loads.

In order to use a once-a-month load for engineering and administration, new peak service criteria and corresponding load objectives must be determined. For small step-by-step offices, peak service criteria have been selected for line finders, connectors, selectors, and only-route trunk groups.

Peak service criteria are determined by calculating the expected blocking or delay that would occur in offices if the once-a-month loads were offered to the amount of equipment required by time-consistent practices. Continuous collection of hourly traffic loads in the study phase of SONDS provided the data needed for this comparison. The following are the objectives for selecting peak service criteria:

1. Provide overall service that is at least as good as that resulting from use of the current time-consistent criteria
2. Provide better service during peak load periods than results from using the current time-consistent criteria
3. Provide service to customers that is independent of the size of the serving office or equipment group
4. Require no more total equipment in the Bell System than is in current use.

Having resolved by field study the EVE blocking objectives for the various components relative to their present TCBH blocking objectives, it is a simple matter to extend the existing capacity tables to the new blocking values and calculate equivalent tables to replace existing TCBH tables. The existence of these extreme value capacity tables is necessary to apply EVE techniques to central office engineering, as discussed in the next section.

VI. TRAFFIC ENGINEERING USING EVE

In the traffic engineering of step-by-step offices, it is necessary to distribute the forecasted originating, local, and incoming traffic loads and to provide equipment quantities to meet objective service criteria.

Customers are associated with a group of line finders that return dial tone and connect the customer to a first selector, which will respond to the first dial pulses. A selector steps vertically in response to each digit (0–9) dialed. It then searches horizontally up to ten terminals for an idle path to a following stage of selectors or for an outgoing trunk. Successive selector stages respond to successive dialed

digits to direct traffic toward outgoing trunks or intraoffice destinations. In the last switching stage a connector completes the call to the customer line corresponding to the telephone number. A connector responds vertically and horizontally in response to the last two numbers dialed. Incoming traffic from other offices is served by incoming selectors associated with incoming trunks. The number of selector switching stages and how the traffic mixes prior to switching to the customer's connector group depend on the numbering plan and size of the office.

To provide measurements for each part of the traffic flow would be too expensive. Using time-consistent practices it has been customary to estimate unmeasured traffic by adding and subtracting traffic measured in other parts of the network. With EVE this is not feasible because the peaks in the various parts of the traffic flow may not occur at the same time.

There are two relatively simple answers to estimating equivalent time-consistent traffic loads, which then may be apportioned as is done in existing step-by-step practices. Both depend to a degree on empirical relationships observed in small offices and relate to the candidate busy hour concept.

The first approach is based on analysis of field data, as was done to determine $h = 6$ of Table IV. In this case, the best fit between the daily peak-hourly values and the TCBH values is for $h = 2.5$. The mean and the standard deviation of the normalized variate for $h = 2.5$ are 0.725 and 0.790, respectively.[7] Using these values in eqs. (11) and (12) we calculate the TCBH equivalent load as:

$$\mu_{\mathrm{TCBH}} = \bar{x} - 0.918\bar{s}, \tag{17}$$

where $\bar{x}$ and $\bar{s}$ are defined in eqs. 4 and 5.

A second method—more appropriate, we think, for estimating the TCBH mean value—is the use of the SONDS peak traffic capacity tables. These tables were obtained by deriving the equivalence over many offices between time-consistent busy hour service criteria and peak service criteria. Traffic capacity tables are available in the SONDS computer for both the peak loads and equivalent time-consistent loads. Using the load values in these tables, we can convert an observed peak load to a time-consistent load.

Table V shows an example comparing the estimates of time-consistent loads derived from once-a-month values with actual measured time-consistent loads for a typical mix of traffic. Both the normal-to-the-2.5 and the capacity table methods give satisfactory results. The table method is preferred, however, because it will come closer to correcting the unmeasured traffic to SONDS objective service levels.

Table V—Intraoffice* traffic estimates (in CCS)

	Once-a-Month Actual	Time-Consistent Estimates		TCBH Actual
		h=2.5	Table	
Terminating	2789	2503	2418	2512
Incoming	1961	1737	1670	1757
Local		766	748	755

* Intraoffice = Terminating-Incoming

VII. SONDS DIAL TONE DELAY ANALYSIS

For SONDS offices, a dial tone delay service index is required for peak measurements, and it should be able to be compared directly with non-SONDS offices using TCBH measurements. The following paragraphs describe the procedures used to make this conversion.

In current practice the service index relates monthly average service results to objective capacity levels. SONDS introduces the concepts of the monthly Average Measured Peak Service (AMPS)[8] as the average of the peak values of the Dial Tone Delay (DTD) results for all the line finder groups in a loading division. An AMPS objective is chosen for the peak service criterion, thereby including the measured volatility of loads within the service objective.

The coefficient of variation, CV, of the current once-a-month load is found from eq. 16 to be:

$$CV = \bar{s}/\bar{x} = (x_{20} - \bar{x})/1.74\bar{x}. \tag{18}$$

An empirical relation between CV and load has been determined.[8] The CV of the capacity load is estimated from that of the measured load by:

$$CV_{\mathrm{CAP}} = CV\left(\frac{x_{20}}{x_{\mathrm{CAP}}}\right)^{0.432}. \tag{19}$$

The value x_{CAP} is the x_{20} value at capacity. Given the CV_{CAP} and the number of line finders per group, the AMPS objective can be selected from Table VI.

These peak values of AMPS and AMPS objectives can be converted[8] to DTD TCBH equivalents from the following empirical relationship derived from data studies.

$$\frac{TCBH\ DTD}{TCBH\ DTD_{\mathrm{OBJ}}} = 0.85\left(\frac{AMPS}{AMPS_{\mathrm{OBJ}}}\right) + 0.17\left(\frac{AMPS}{AMPS_{\mathrm{OBJ}}}\right)^2. \tag{20}$$

SONDS uses this equation to convert the measured AMPS values to equivalent TCBH DTD values. These values of TCBH DTD can be entered directly into current, accepted index tables.

Table VI—AMPS objective (percent) coefficient of variation of daily peak loads at capacity

		0	0.05	0.10	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
Number of line finders per group	1	9.0	8.3	7.6	7.2	6.7	6.3	5.9	5.6	5.3	5.1	4.8
	2	9.2	7.9	6.9	6.1	5.5	5.0	4.6	4.2	3.9	3.7	3.5
	3	9.3	7.8	6.6	5.6	4.9	4.4	3.9	3.6	3.3	3.1	2.9
	4	9.4	7.4	6.0	5.0	4.3	3.8	3.4	3.1	2.8	2.7	2.5
	5	9.6	7.4	5.8	4.8	4.0	3.5	3.1	2.9	2.7	2.5	2.4
	6	9.7	7.1	5.4	4.4	3.7	3.2	2.8	2.6	2.4	2.3	2.2
	7	9.7	7.0	5.3	4.2	3.5	3.0	2.7	2.5	2.3	2.2	2.1
	8	9.9	6.9	5.1	4.0	3.3	2.9	2.6	2.4	2.2	2.2	2.1
	9	9.9	6.8	4.9	3.8	3.2	2.8	2.5	2.3	2.2	2.1	2.1
	10	10.0	6.6	4.8	3.7	3.0	2.6	2.4	2.3	2.2	2.1	2.0
	11	10.1	6.6	4.7	3.6	3.0	2.6	2.4	2.2	2.1	2.1	2.1
	12	10.2	6.5	4.5	3.5	2.9	2.5	2.3	2.2	2.1	2.1	2.1
	13	10.3	6.4	4.4	3.3	2.8	2.5	2.3	2.2	2.1	2.1	2.1
	14	10.3	6.4	4.3	3.3	2.7	2.4	2.3	2.2	2.1	2.1	2.1
	15	10.4	6.3	4.3	3.2	2.7	2.4	2.3	2.2	2.2	2.1	2.1
	16	10.5	6.2	4.2	3.2	2.7	2.4	2.3	2.2	2.2	2.2	2.2
	17	10.5	6.1	4.1	3.1	2.6	2.4	2.3	2.2	2.2	2.2	2.2
	18	10.6	6.0	4.0	3.0	2.6	2.3	2.2	2.2	2.2	2.2	2.2
	19	10.7	6.0	3.9	3.0	2.6	2.4	2.3	2.2	2.2	2.2	2.2
	20	10.7	6.0	3.9	3.0	2.5	2.4	2.3	2.2	2.2	2.3	2.2
	21	10.8	5.9	3.8	2.9	2.5	2.3	2.3	2.2	2.2	2.3	2.3
	22	10.9	5.8	3.8	2.9	2.5	2.3	2.3	2.3	2.3	2.3	2.3
	23	10.9	5.8	3.7	2.8	2.5	2.3	2.3	2.3	2.3	2.3	2.4
	24	11.0	5.8	3.7	2.8	2.5	2.4	2.3	2.3	2.3	2.4	2.4
	25	11.1	5.7	3.6	2.8	2.5	2.4	2.3	2.3	2.4	2.4	2.4
	26	11.1	5.7	3.6	2.8	2.5	2.4	2.3	2.3	2.4	2.4	2.4
	27	11.2	5.7	3.6	2.8	2.5	2.4	2.4	2.3	2.4	2.4	2.5
	28	11.2	5.6	3.5	2.7	2.5	2.4	2.4	2.3	2.4	2.5	2.5
	29	11.3	5.6	3.5	2.7	2.5	2.4	2.4	2.4	2.5	2.5	2.5
	30	11.3	5.6	3.5	2.7	2.5	2.4	2.4	2.4	2.5	2.5	2.6

VIII. SONDS LINE ASSIGNMENT AND LOAD BALANCE

Line assignment is an important task of the network administrator and a load balance index is part of TNDS. The procedures for both of these are well established but are based on TCBH techniques. Farel[10] derived new calculations to use the once-a-month loads obtained by SONDS. Consistent with the previous techniques, the SONDS line assignment recommendations are smoothed over time so that stable line assignment procedures are obtained. Also the load balance index reflects the level of load carried by the office; in particular, it recognizes that bad load balance does not produce bad service in a lightly loaded office.

Farel's line assignment algorithm[10] is based on the load deviation of each line finder group from the average load of the loading division, as follows:

$$\bar{u}_i(t) = w_i(t)F \tag{21}$$

and

$$w_i(t) = a[\bar{u}_i(t-1) + \lambda(k_i - \bar{k})] + (1-a)y_i(t), \tag{22}$$

where:

$\bar{u}_i(t)$ is the smoothed once-a-month load deviation for the ith line finder group for month t,

$\bar{u}_i(t-1)$ is the same quantity for the previous month,

λ is the CCS/MS derived for the current month by dividing the sum of the once-a-month loads in hundreds of call seconds (CCS) for the loading division by the sum of the main stations (MS) in the loading division,

k_i is the change in the number of main stations from last month for the ith line finder group,

$\bar{k}$ is the average of k_i,

$y_i(t)$ is the deviation of the once-a-month load for the ith line finder group from the average of the loading division,

a is a constant that weights the previous month's data to those of the current month,

and

$$F = 1 - (N - 3)(.034\bar{z})^2/\sum w_i^2, \tag{23}$$

where:

N is the number of line finder groups in the loading division,

$\bar{z}$ is the loading division average of the line finder group once-a-month loads,

w_i is an intermediate variable, given by eq. (22).

The line assignment recommendation for the ith line finder is:

$$q_i(t) = -\bar{\mu}_i(t)/\lambda \tag{24}$$

rounded to the nearest integer. A positive value of $q_i(t)$ indicates that lines should be added to line finder group i.

Examination of eq. (22) shows that the inclusion of the k term accounts for the line assignment activity (connects and disconnects) of the previous month.

The value of a has been set at 0.32 from data studies. This means that 68 percent of the weight of the current line assignment recommendation is based on the current month's data; 32 percent of the weight is based on history. This leads to consistent month-by-month line assignment recommendations.

The quantity F of eq. (23) is the James-Stein factor and is a measure of the confidence to be placed on the calculations based on statistical volatility of the data. It is always less than unity and typically runs about 0.9.

The same quantities used to generate the line assignment recommendations are also used in the load balance index calculation.[11] First the nonstatistical deviation of the current month's loads $D(t)$ is calculated as:

$$D(t) = \sum w_i^2/N - (.034\bar{z})^2(N - 1)/N. \tag{25}$$

This current month's value is used with that of the previous month $D(t-1)$ to derive a smoothed value L by a constant b as follows:

$$L = bD(t-1) + (1-b)D(t). \tag{26}$$

The value of b has been set at 0.7 from data studies. This means that 30 percent of the weight of the smoothed deviation is based on the current month's data; 70 percent is based on history.

Finally, an imbalance factor is derived as:

$$(V/M)_{\text{OAM}} = L/\bar{z}. \tag{27}$$

While this factor is a direct measure of the load imbalance, the calculation of the index has to account for the level of load being carried by the office, called percent capacity C. It is calculated as:

$$C = \bar{z}/z_{\text{CAP}}. \tag{28}$$

The value of z_{CAP} is obtained from the SONDS capacity table for the number of line finders in the line finder group.

The imbalance factor and the percent capacity are used to derive the load balance index, as shown in Table VII. This table shows that a given imbalance factor results in a lower index as the carried load approaches capacity. Therefore, the index is a measure of service as seen by the subscribers, while the imbalance factor is a message to the network administrator to bring the office into balance as the load builds up.

The imbalance factor is calculated for each loading division because the lines are assigned by loading division. The office load balance index is derived by combining the imbalance factors of all the loading

Table VII—SONDS load balance index

Imbalance Factor V/M_{OAM}	Percent Capacity							
	≥96	95–86	85–76	75–66	65–56	55–46	45–36	≤35
0–0.8	100	100	100	100	100	100	100	100
0.8–1.3	99	99	100	100	100	100	100	100
1.3–1.7	98	99	99	100	100	100	100	100
1.7–2.1	97	98	99	99	100	100	100	100
2.1–2.6	96	97	98	99	100	100	100	100
2.6–3.3	94	96	97	98	99	100	100	100
3.3–4	92	95	96	98	99	99	100	100
4–5	90	93	95	97	98	99	100	100
5–6	85	92	94	96	98	99	99	100
6–7	75	90	93	95	97	98	99	99
7–8	63	87	91	94	96	98	99	99
8–10	45	82	90	93	96	97	98	99
10–13	18	68	87	92	95	97	98	99
13–16	11	57	84	91	94	96	97	98
16–25	1	20	64	87	92	95	97	98
25–33	0	11	48	83	91	94	96	97
33–50	0	6	28	74	89	93	95	97
≥50	0	1	14	61	86	91	94	96

divisions according to the relative number of main stations within each loading division.

IX. ANALYSIS OF EVE DATA

One of the advantages of EVE techniques is that one can determine if a set of data points follows an extreme value distribution by establishing a set of quality control limits. It is seen that the data points of Fig. 3 violate their quality control limits. When nine low-side outliers were removed, the remaining data points, shown in Fig. 4, fall within their (updated) quality control limits.

This feature is used in SONDS in a powerful data validation test to determine if a new data point is an acceptable member of an established distribution. But first one has to establish the reference distribution for each usage measurement.

9.1 Start-up tests

The reference distribution is established in SONDS by accepting 20 days of peak data with no screening. At the end of this time, called the start-up period, three low-side, followed by two high-side, outlier tests are applied to the data. These are called censored outlier tests because the contribution of the data point being tested is removed from the estimates of the mean and the variance. This is done so that a contaminated data point will not be accepted because of its influence on the test conditions.

The following equations are given by Friedman[8] for estimating the mean and the variance of the equal "candidate busy hour" load distribution from the censored sample set of peak data points via the moments method. The censored mean is

$$\hat{\mu} = \bar{x}_{n-1} - \frac{\hat{\sigma}}{n-1}\,\{1.267n - E[y_{(n)}]\} \tag{29}$$

and the censored variance is

$$\hat{\sigma}^2 = \bar{s}^2_{n-1}\,(n-2)/(0.416(n-1) - \frac{n}{n-1}\,\{E[y^2_{(n)}] - 2.534E[y_{(n)}] + 1.605\}) \tag{30}$$

where

$\bar{x}_{n-1}$ is the sample mean and $\bar{s}^2_{n-1}$ is the sample variance of the remaining $n - 1$ data points as each highest or lowest data point is removed

and

the reduced variate $y = (x - \mu)/\sigma$.

Table VIII—Low-side and high-side censored means and variances

n	Low-Side		High-Side	
	$E[y_1]$	$E[y_1^2]$	$E[y_n]$	$E[y_n^2]$
17	—	—	2.515	6.50
18	0.182	0.114	2.535	6.60
19	0.168	0.108	2.554	6.70
20	0.156	0.102	2.572	6.79

$E[y_{(n)}]$ is the expected value of the largest of the n data points, each of which has been selected as the largest of six samples drawn from a normal distribution with zero mean and unity variance. $E[y_{(n)}^2]$ is the expected value of the largest square of n such data points. $E[y_{(1)}]$ and $E[y_{(1)}^2]$ are the equivalent values for the lowest of n data points. Values for both sets of parameters as used in SONDS are listed in Table VIII as given by Friedman.[8]

Using the censored mean and variance of eqs. (29) and (30) with $E[y_{(1)}]$ and $E[y_{(1)}^2]$, we compute $F(x_L)$ from eq. (8). We then compute the low threshold L as:

$$L = 1 - [1 - F(x_L)]^n \leq 0.06. \tag{31}$$

If the value of L is less than 0.06, the data point is rejected and the next lowest data point is tested. This value of the threshold will reject six percent of valid lowest-of-twenty data points.

Using the censored mean and variance of eqs. (29) and (30) with $E[y_{(n)}]$ and $E[y_{(n)}^2]$, we compute $F\ (x_H)$ from eq. (8). We then compute the high threshold H as:

$$H = 1 - [F(x_H)]^n \leq 0.01. \tag{32}$$

If the value of H is less than 0.01, the data point is rejected and the next highest one is tested. This value of the threshold will reject one percent of valid highest-of-twenty data points.

If either three low-side data points or two high-side data points fail the threshold tests, the complete set of start-up data is rejected and a new start-up interval is automatically initiated. Restarts almost completely disappeared after the normal-to-the-sixth replaced the Gumbel as the assumed distribution.

If both thresholds are passed, the particular measurement is considered operational and a different set of data validation tests are applied, as described in the next section.

9.2 Operational tests

When the start-up data set for a particular measurement is accepted, its mean and variance are used for the initial operational data valida-

tion tests. The values of the mean and the variance used for subsequent operational tests are continually updated as new data are accepted. The failures of certain of these operational tests will cause an exception report to be issued. Other test failures will increment a counter, which may or may not result in an exception report. If a given day's data contain too many exception conditions on individual measurements, the whole set of daily data is rejected and an exception report is issued.

The following are the operational tests applied to EVE data.

9.2.1 *Upper physical bounds test*

The received data should not be greater than 36 CCS times the number of components. (This test is also applied to non-EVE usage measurements.)

9.2.2 *Lower physical bounds test*

The received value must be greater than zero. A zero value for the daily peak usage measurement of a group of components to be subjected to EVE analysis would indicate a trouble condition or a mistake in measurement assignment. (Zero is allowed for both non-EVE usage measurements and peg count measurements.)

9.2.3 *Once-a-month load exceedance test*

This test compares each new data point with the once-a-month load computed from eq. (16) with the current values of mean and variance. A data point higher than that value will increment an up-down counter; a data point lower than that value will decrement the counter (but not below zero). An exception report is issued at the count of three. In this way traffic trends or bad data can be detected.

9.2.4 *Coefficient of variation test*

If a register or piece of equipment is inoperative, the coefficient of variation of the moving average of the mean and the variance falls below the threshold value of 0.025, and an exception report is issued.

9.2.5 *High and low outlier tests*

Before a new EVE data point is accepted, it must pass both a low and high outlier test. The current values of the mean and the variance are used to compute $F(x_i)$ from eq. (8).

The low threshold L is computed as:

$$L = 1 - [1 - F(x_i)]^{20} \leq 0.06. \tag{33}$$

If the value of L is less than 0.06, the new data point is rejected and it is not used to update the mean or the variance. This threshold value will reject only about one in 300 valid daily peak values.

The high threshold H is computed as:

$$H = 1 - [F(x_i)]^{20} \leq 0.01. \quad (34)$$

If the value of H is less than 0.01, the new data point is rejected and not used for updating. This threshold value will reject only about one in 2000 valid daily peak values.

By solving eqs. (33) and (34) numerically and using eqs. (13) and (14), the acceptable range for a new data point in terms of the current values of the mean and standard deviation can be found as:

$$\bar{x} - 2.442\bar{s} \leq x_i \leq \bar{x} + 3.889\bar{s}. \quad (35)$$

X. SYSTEM CONFIGURATION

A block diagram of SONDS is shown in Fig. 5. The most prominent feature is the single, central computer for the Bell System. The source of the traffic data is a data collection unit in each step-by-step office. Output reports are delivered to the user by shared access to a local reports terminal. The user controls the system by means of the user interface. All of these elements are connected together by the switched network.[12]

One of the advantages of the system is its use of the low-cost switched network instead of dedicated data links for the data collection and report distribution. The use of the switched network and computer resources is optimized by middle-of-the-night polling, followed by early morning reports transmission. The scheduled reports are available to the user at the start of business on the due date, and often contain results as recent as the previous day.

Consolidating the software in a single computer led to reduced development costs and timely revisions to correct bugs or to incorporate enhancements. The time-shared nature of the computer imple-

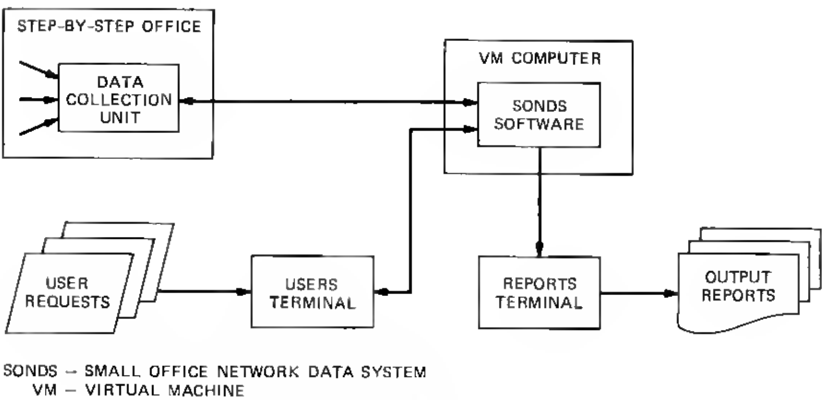

Fig. 5—Block diagram of SONDS.

mentation made it natural to put the user manual on-line. This not only reduced the cost of documentation but made it more timely, user-oriented, and easily accessible. This type of documentation has reduced costs further by precluding the need for any formal training of the (many and casual) users.

Individual sections of the user manual can be obtained on-line. Also, the entire manual or any of its sections can be printed off-line and mailed to the user as the result of an on-line request. The date of the user manual is the date of its most recent change and the user is notified of this each time he or she logs onto the computer. The table of contents contains the date of tbe most recent change of each section. Each page contains the date of the most recent change incorporated thereon and the change itself is noted by marks in the margin.

To enhance the fully mechanized and less paper aspects of the system, the computer contains the capacity tables for the components to be engineered by use of the system's data. This allows the system to produce fully analyzed results, requiring no further manual manipulations by the user.

The user interface is the key to making the system easy to operate and completely under user control. The most important function of this interface is entering the office information that SONDS needs to produce the output reports. This information includes the component names that SONDS presents in the output reports. It also includes the number of the components that SONDS needs to access the capacity tables for those components. Further, SONDS must know about the units of measurement to convert the received data to the proper units: CCS for usage, units for peg counts, and percent for dial tone delays.

The interface contains a series of prompts that help the user enter the information. Out-of-bound values (probably typograpbical errors) are detected immediately and the user can correct them at this time or add them later if the proper value is unkown. Whenever the user does not know what to do next, typing a question mark will produce a list of acceptable inputs.

Another application of the user interface is to request previously generated reports that were either misplaced or not received because of a reports terminal problem. Typically, the most recent three or four scheduled reports are kept in this on-line archive; the (daily) exception reports are kept for two weeks. Also, the user can demand a current version of any of the scheduled reports. He will receive a partial version of the next scheduled report. This request, however, will not interfere with the regular generation of the scheduled report.

Perhaps the most powerful use of the interface is to command the system to go into hourly data collection. Without interfering with the regular nightly polling and report generation, the computer will initiate

a data poll each hour between 6 a.m. and midnight, or as specified by the user. Thus, time-consistent hourly data can be generated. This feature is needed for limited use to verify data for a newly equipped office, to troubleshoot a data problem, or to conduct special studies.

10.1 Data collection unit

To obtain the desired data from the step-by-step offices, SONDS has specified a set of features, some of which are optional, for a Data Collection Unit (DCU). The three vendors and the four models of DCUs used on SONDS are Western Electric (PDT2A) (13), Tele-Sciences (TE450), and Alston (Model 565 and 566). Each microcomputer-controlled data terminal is connected to a data set that has an automatic answer capability. This is used for dial-up access by the SONDS computer to obtain the measurement data. This dial-up access can also be used by maintenance people for remotely running diagnostics, as specified by the vendor.

The DCU collects both peg count and usage data, the latter for either one-second or ten-second scan rates.

The output register set for each measurement contains three registers. One is the active register, whicb accumulates the counts during the measurement hour. At the end of each hour the active registers are reset automatically and the contents are made available to the previous-hour and long-term registers.

For each measurement, the new hourly value from the active register overwrites the value contained in the previous-hour register where it is retained for the next hour. The new hourly value also goes to the long-term register, which has been specified to be either a peak or an accumulative register. If the long-term register is peak, the new hourly value is compared to that stored in the peak long-term register and the higher of the two is retained in the peak long-term register. If the long-term register is accumulative, the new hourly value is added to its contents.

The long-term registers are reset after reading by the SONDS computer during the nightly poll. The values obtained from the long-term registers are either the highest hourly value experienced since the last resetting or the accumulations of the hourly values since the last resetting.

The DCU reads out the register contents on command from the SONDS computer. During the nightly poll, the computer requests the contents of both the previous-hour and long-term registers and then the computer resets the long-term registers. For the collection of hourly data during the daytime, the computer only requests the contents of the previous-hour registers and disconnects without any resetting. The SONDS computer will accept up to 200 register readings in response to a given request.

Consideration of mid-train engineering and the possible use of SONDS data for the pre-installation engineering of an electronic replacement office has led to the definition of two optional features. One is called the peak of the sum. What is desired here is the (24-hour) peak value of a measurement that itself is the sum of other measurements. Since the individual measurements might peak at different hours, the sum of the (24-hour) peaks will, in general, not be equal to the (24-hour) peak of the sum of the bourly values. Therefore, the hourly summing and (24-hour) peaking must be done in the DCU. The output is available in one of the regular peak long-term registers.

The other optional feature is called the peak of the difference. The desired result is the peak value of a measurement that itself is the difference between two sets of other measurements. Again, the peak of the difference is, in general, not equal to the difference of the peaks. The output is available in one of the regular peak long-term registers.

Besides these measurement features, the DCU satisifes several other SONDS requirements. Each response has a specific format that the SONDS computer expects. It includes the DCU time of day, which is resettable by the SONDS computer, and a status word to indicate several conditions within the terminal that might make the data suspect and that would aid in the trouble location of a DCU problem. The response also contains a checksum value of the response itself. The SONDS computer calculates the checksum of the received data and compares it to the received checksum. In this way the SONDS computer can be assured that it has an accurate copy of the register readings as they exist in the DCU.

10.2 Data calculations

To generate the several types of output reports, SONDS has four different types of data analysis. In entering the office information, the user selects one type of analysis for each measurement.

10.2.1 EVE analysis

EVE analysis is used for the usage measurements in the main part of the SONDS monthly report and for selected measurements in the miscellaneous section of that report.

As seen in eq. (16), the once-a-month load is calculated from the mean and the standard deviation (the square root of the variance) of the daily peaks. Therefore, it is not necessary to retain all the received values; only an estimate of the mean and the variance is required. This is accomplished in SONDS as a continual updating of these values, once they are derived at the end of the start-up period.

The equations for this updating are as follows:

$$\bar{x}_i = \mathrm{p}x_i + (1 - \mathrm{p})\bar{x}_{i-1} \tag{36}$$

$$v_i = p(x_i - \bar{x}_i)^2 + (1 - p)v_{i-1}, \tag{37}$$

where

x_i is received value for the ith day,

$\bar{x}_i$ is the mean at the ith day,

$\bar{x}_{i-1}$ is the mean from the previous day,

v_i is the variance at the ith day,

v_{i-1} is the variance from the previous day,

and

p is a constant.

This updating method reduces the size of the database. Only two values need to be stored for each measurement; no raw data are retained. It gives an updated value for the mean and the variance, which is the exponential weighting of all the previously received data. The value of p was selected as 0.095 to make the average age of the weighted data the same as if the latest month of data itself had been used. One effect of this calculation is that 87 percent of the weight of the current estimate is derived from data received during the most recent month; 13 percent of the weight is due to data received in prior months. This approach, besides minimizing data storage requirements, is also compatible with current TCBH practice of reporting monthly results.

Another advantage of this updating technique described in eqs. (36) and (37) is that it makes the calculation robust for (occasional) losses of data. For any day that data are not obtained, the estimate is not updated, that is, the estimate is stale by that amount. The exponential weighting of previous data reduces this staleness, and the effect of a lost data point is soon eliminated.

10.2.2 PA analysis

Periodic Average (PA) analysis is used for dial tone delays and for those usage measurements in the Miscellaneous section of the SONDS monthly report that do not lend themselves to the EVE analysis.

The equation to update the estimate of the mean is:

$$\bar{x}_j = x_j/j + \bar{x}_{j-1}(j - 1)/j, \tag{38}$$

where

j is the jth day during the current period.

This analysis type is called Periodic Average because it calculates the true, straight-line average during the period of the monthly report; the value is reset when the monthly report is generated. PA also retains the single largest value received during the calculation period.

10.2.3 SMA analysis

The third type of analysis is called the Service Month-to-date Accumulation (SMA). This analysis type is used for the Automatic Number Identification (ANI) measurements needed for the Network Switching Performance Measurement Plan (NSPMP). It basically gives the month-to-date summation of the daily values of the (24-hour) accumulative long-term registers. Similar to the PA analysis, the SMA analysis also produces the periodic average and the single highest received value.

10.2.4 TOPC analysis

The fourth type of analysis is called Total Office Originating Peg Count (TOPC). It is specified during the entry of the office information to be applied to originating PC measurements. These measurements are stored for the generation of the (calendar) month TOPC report in which certain weekday and weekend ratios and averages are calculated. Also certain adjustments are made for the loss of individual days of data.

10.3 Output reports

As part of its goal to be a fully mechanized system, SONDS automatically calculates, generates, and transmits several output reports. The monthly report, the (monthly) Total Office Originating Peg Count Report and the weekly report are regularly scheduled. Exception reports are issued daily as needed. Only the hourly data report depends on user initiation (to request hourly polling). It is generated and transmitted daily to give the previous day's polling results.

10.3.1 Monthly report

The heading of the monthly report gives the office name and the date of the report, which is typically the 22nd of the month. The report heading also gives the date that the office information file was last updated; this is a message to the user because his/her monthly updates of office component counts can affect the numbers presented on the report.

Another part of the header information is the number of days of valid EVE data. Although SONDS collects data seven days a week, EVE data are only used for five days. Typically, these are the weekdays, but the user can include Saturday, Sunday, or both in the five days if weekends are high traffic days in a particular office. There are only 20 to 23 EVE days in a month and the number of valid days indicates the weight of the EVE results due to data collected during the current month. Fifty percent of the weight occurs in the current month if seven days of valid EVE data are obtained. Data for measurements

with fewer than seven days in the current month are shown on the monthly report, but they are flagged with a ?. Incidentally, if a measurement is in start-up, it is flagged with an asterisk.

The first part of the Originating Results section of the monthly report contains information on the Line Finder Groups (LFGs). This includes the items discussed in prior sections, such as the current load as percent capacity, main station assignment guide, and maximum value of Dial Tone Delay (DTD), all on a per-LFG basis. Also there are total office results in terms of the Dial Line Index (DLI) and Load Balance Index (LBI).

The Originating Results section of the monthly report also contains an outgoing trunk group section. Certain columns containing information not previously discussed were specifically designed for the trunking people. One such column gives the number of circuits required to carry the measured load at the once-a-month blocking objective; this quantity is equivalent to one of the main outputs of the Trunk Servicing System (TSS). The second column can be used as the TSS (TCBH) study period load; it is derived from the proper TCBH trunk capacity table, depending on the trunk accessing arrangement, for the number-of-trunks required specified in the previous column. SONDS retains these study period values for the most recent 12 montbs and each month prints the highest value in the third column. This is the base period load, which can be the input to the Trunk Forecasting System (TFS). The fourth column prints the month when this highest load occurred.

The Terminating Results section of the monthly report contains similar information for selectors, connectors, and incoming trunks.

The Miscellaneous Results section contains any other measurements the user has decided to add to the system. The user can select peg count or usage measurements, peak or (24-hour) accumulating registers, and has a choice of EVE, PA, or SMA analysis types. One of the main uses of this section has been to obtain the ANI counts for NSPMP reporting. Another use has been to obtain the results of the summing and differencing registers to obtain data for pre-installation engineering of an electronic replacement office. Of course, the user could also choose to obtain Last Trunk Usage (LTU) or Subscriber Line Usage (SLU) measurements.

10.3.2 Weekly report

The Weekly report is scheduled for Monday mornings; it gives daily results for seven days up to the previous Friday. The main part of the report gives the "always busy" data obtained by the SONDS computer during the nightly polling by reading the previous-hour registers of the data collection unit. Integer multiples of 36 CCS in these data are

indicative of busied-out or stuck components. The report also contains the daily peak values of dial tone delays so that the user can determine if high values of dial-tone delays were caused by maintenance conditions indicated by the always-busy data. The report contains the daily values of total office originating peg counts to make these data available to the Division of Revenues people before the end of the (calendar) month. The report contains the month-to-date results of the DTD (TCBH) and NSPMP measurements so the user can monitor the reportable service results during the month.

10.3.3 Exception reports

Exception Reports (ER) are issued daily as needed when abnormal conditions are experienced during the night's poll. These conditions might be a failure to connect (after four attempts), error bits set in the status word, and failed data validation tests. These ERs are short messages giving the ER code, the descriptive name, and the received data, if any. There are some 52 different types of ERs and each has a one- to three-page description in the SONDS User Manual to help the user locate the problem. Also, the user can access the SONDS computer to obtain a list of the ERs for the past two weeks.

10.3.4 Total Office Originating Peg Count report

The Total Office Originating Peg Count report has a calendar-like format showing the daily values with weekday values separated from weekend values. Weekday-to-weekend ratios are calculated as well as values of average business days and calendar days. The computer contains algorithms to adjust for isolated days of missing data and it computes adjusted averages.

10.3.5 Hourly data report

Hourly data reports are generated only when hourly polling has been requested by a user. The reports are received in the morning and contain up to 16 hours of polling results of the prior day. This produces time-consistent hourly data, including dial tone delays, to track the flow of traffic into, out of, and across the office. For each hour, the report also contains a traffic balance calculation to determine if permanent signals are inflating the measured value of line finder usage.

10.3.6 Retention periods

As mentioned previously, SONDS retains on-line for user access two weeks of exception reports and the last three or four reports of other types. The user can also receive on-line a demand version of any

of the reports, which can supply data as recent as that of the previous day.

XI. COMPUTER AND SOFTWARE DESCRIPTION

The SONDS software consists of approximately 260 programs residing on an AT&T Information Systems computer located at the Corporate Computer Center in Piscataway, N.J. SONDS shares with many other corporate users an Amdahl 470/V8 processor under the Virtual Machine/Conversational Monitor System (VM/CMS) operating system.

11.1 VM/CMS computer facility

VM/CMS manages the resources of a real-time computing system in such a way as to make them available to many users at the same time. Through VM/CMS, each SONDS user has a virtual machine composed of a virtual system console (a terminal), virtual CPU, virtual storage (disk), and virtual channels and Input/Output (I/O) devices.

The VM/CMS Auto-dial Facility allows application programs to initiate a direct distance dialing connection to a remote terminal properly equipped with an auto-answer feature. SONDS uses the Auto-dial Facility in two ways: It extracts data at a 300-baud rate from the data collection unit in each SONDS office by nightly polling, and it prints output reports on users' printing terminals at the rate of either 300 or 1200 baud. This report distribution process is faster and less expensive than mailing reports directly and can make use of terminal equipment outside normal working hours.

11.2 SONDS software

Each of the major modules of the SONDS software is described in the following sections.

11.2.1 User interface

The user enters and maintains office information on the user interface. It can also be used to change the polling mode of the office: start (nightly) polling, stop polling, and schedule hourly polling.

The user accesses the Virtual Machine (VM) system over the switched network via any ASCII-compatible terminal at either 300 or 1200 baud. After entering a user identification (userid) and a password, the user is connected to the SONDS user interface. The system gives any short news items that the user should know, and the date of the last update to the User Manual. Whenever the user doesn't know what to do next in the interactive dialogue with the system, he or she can enter a question mark. The system will respond with a list of appropriate entries.

11.2.2 Users' database

The office information file entered by the user via the user interface is stored in the users' database. Also the traffic information obtained from the data collection unit is stored here. All the data are stored on an office-by-office basis. Some administrative reports, however, do combine information on how SONDS is working for all the offices on a userid and all the offices for a company.

11.2.3 System files

The system files primarily contain office status information to determine which activities should be started and when it is appropriate to start them. They are updated as the scheduled activities are performed. They thereby always contain the current information and can be queried by the SONDS Support Team to provide a current report of system status.

11.2.4 Master scheduler

The master scheduler controls four main activities: nightly polling, hourly polling, report generation, and report transmission.

Nightly polling occurs seven days a week, including holidays, during low traffic periods, usually between midnight and 6:00 a.m. local office time. The polling load is distributed over as many hours as necessary, taking advantage of the different time zones across the country.

Hourly polling may be requested by the user through the user interface for as many as six days of polling. The user specifies the earliest hour and the latest hour for polling, between 6 a.m. and midnight. Hourly polling can be started on the next clock hour. Hourly polling increases SONDS costs and charges, which in turn have prevented abuse of this feature.

The master scheduler activates the report generator and report transmitter processors according to the schedule for each of the five output report types, described in Section 10.3. The master scheduler also establishes an efficient order of nightly polling, hourly polling, report generation, and report transmission by considering the number of Automatic Calling Units (ACUs), the number of offices per time zone, estimated time to perform each function, etc.

The master scheduler runs each hour, usually starting at one minute past the hour and running for 15 seconds to 2 minutes. In addition, one of the major objectives of the master scheduler is to invoke "batch jobs" as needed for the poller, the report generator, and the report transmitter. Once initiated, these batch jobs will run on their own without master scheduler intervention.

11.2.5 Poller

The principal poller is activated during the night to place a call over the switched network to the data collection unit using the ACUs. The data as well as the status information of the DCU are sent to the users' database and the system files are updated. The action of the hourly poller is similar except that it is activated during daytime hours.

11.2.6 Report generator and transmitter

When the report generator is activated by the master scheduler, it obtains the traffic data from the users' database, applies the necessary algorithms to calculate the numbers that appear on the report, and stores the report data. The report transmitter then places a call over the switched network to the user-specified reports terminal via the ACUs, retrieves and formats the report data, and transmits the report to the users' terminals over the switched network.

11.2.7 Map Management System

The PDT2A data collection unit contains a software map to cross connect the input leads to the output registers.[13] This software map is created by the Interactive Map Assembly Program (IMAP), which exists on a time-shared DEC 10 computer at a Western Electric facility. Using the VM batch facility, the Map Management System will accept a user-generated request to place a call over the switched network via the ACUs to access IMAP and obtain the most recent edition of the software map, which is then stored in the users' database for down-loading to the PDT2A via the poller.

11.2.8 Administrative System

The Administrative System is a group of programs that monitor the activities of all the other modules, recording their successes and failures. The system generates several types of detailed logs and summary reports, which are available to the SONDS Support Team to determine the health of the SONDS and VM systems and to identify and resolve problems in a timely manner. The user has access to a (pertinent) subset of these administrative reports via the user interface.

11.2.9 Billing system

Billing for VM/CMS is done once a month and a separate statement is produced for each userid. Each statement includes a detailed listing of the charges for each off-line activity the userid had during the month, and a breakdown of the usage on the userid by accounting information, which identifies the initiator of the on-line activity and

the type of off-line activity. To increase the user's cost consciousness, the user interface prints an approximate charge, upon logoff, for that user session.

XII. ACKNOWLEDGMENTS

The authors are indebted to K. A. Friedman, who was the main analyst throughout this project. He proposed and analyzed the normal-to-the-sixth distribution, the outlier detection tests, and the dial tone delay conversions. He collected and analyzed the vast quantities of field data needed to derive extreme value service criteria. He derived the new engineering tables used in SONDS for all the major components of the step-by-step office. R. A. Farel's work on the use of extreme values for line assignment and load balance, the results of which are published for the first time herein, was a significant contribution towards making SONDS a member of the TNDS family. We are particularly appreciative of the contributions, suggestions, and critiques of J. G. Kappel throughout the project. The experimental version of SONDS, which was used to develop many of the concepts reported herein, was possible due to the talent of the software team led by G. B. Foley and S. Chen and to the dedicated field support of R. R. Plum. The one thousand SONDS offices now operating on the VM computer would not have been possible without the dedicated effort of J. Schwartz and his software development team at AT&T Information Systems. Their generous review of the pertinent parts of this paper is gratefully acknowledged; the significant contribution of K. L. Dziedzic is specifically acknowledged.

REFERENCES

1. S. A. Karlsson, "The Dimensioning of Telephone Traffic Routes for Measured Integrated Peak Traffic," Fifth Int. Teletraffic Cong., New York, June 14–20, 1967, pp. 7–17.
2. S. B. Gershwin, R. V. Laue, and E. Wolman, "Peak-Load Traffic Administration of a Rural Line Concentrator," Seventb Int. Teletraffic Cong., Stockholm, Sweden, June 13–20, 1973, pp. 146/1–7.
3. S. B. Gershwin, R. V. Laue, and E. Wolman, "Peak-Load Traffic Administration of a Rural Multiplexer With Concentration," B.S.T.J., *53,* No. 2 (February 1974), pp. 261–81.
4. R. V. Laue, "Extreme-Value Engineering Designs for Peak Traffic," Bell Lab. Rec., *56,* No. 2 (February 1978), pp. 38–42.
5. E. J. Gumbel, *Statistics of Extremes,* New York: Columbia University Press, 1958.
6. D. H. Barnes, "Extreme Value Engineering of Small Switching Offices," Eighth Int. Teletraffic Congr., Melbourne, Australia, November 10–17, 1976, pp. 242/1–7.
7. D. H. Barnes, "Observations of Extreme Value Statistics in Small Switching Offices," Ninth Int. Teletraffic Cong., Torremolinos, Spain, October 16–24, 1979, 9 pages.
8. K. A. Friedman, "Extreme Value Analysis Techniques," Ninth Int. Teletraffic Congr., Torremolinos, Spain, October 16–24, 1979, 7 pages.
9. L .H. C. Tippett, "On the Extreme Individuals and the Range of Samples Taken from a Normal Population," Biometrika, *17* (1925), pp. 364–87.
10. R. A. Farel, unpublished report.
11. R. A. Farel, unpublished report.

12. G. B. Foley, J. J. O'Connor, and J. Schwsrtz, "SONDS Ssys 'Sound Off' to Small Offices," Bell Lsb. Rec., *58*, No. 10 (November 1980), p. 333–8.
13. J. Arthur Grandle, Jr., and Richard R. Plum, "A Versatile Data-Gathering Tool," Bell Lab. Rec., *57*, No. 8 (September 1979), pp. 227–31.

AUTHORS

Douglas H. Barnes, B.S. (Industrial Engineering), 1938, Lehigb University. Mr. Barnes joined Ohio Bell Telephone Company in 1941 as an Assistant District Traffic Superintendent, where he worked on a variety of traffic assignments including operator force management, traffic staff, service observing, local and toll office traffic engineering, trunk engineering, and division of revenue studies. He came to Bell Lahoratories in 1951 to work on requirements for traffic measurement needs and systems. He was appointed Head of the Traffic Systems Engineering and Analysis Department in 1961. He was responsible for systems engineering of all traffic data collection and processing systems. Some of these are: Traffic Usage Recorders, Traffic Data Recording System, Engineering and Administration Data System, Mechanized Service Ohserving, Force Administration Data System, Subscriber Line Usage System, and Small Office Network Data System. He was also responsihle for various improvements in traffic methods, such as the adaptation of Extreme Value Engineering. He retired from Bell Lahoratories in 1979 and has since heen a Teletraffic Consultant.

John J. O'Connor, B.S.E.E., 1951, M.S.E.E., 1952, Columbia University; Ford Instrument Co., 1952–1954; Electronics Research Lahoratories, C.U., 1954–1958; Hathaway Instrument Co., 1959–1961; Martin Co., 1961–1964; Bellcomm, 1965–1972; Bell Laboratories, 1972—. At Ford Instrument Co. Mr. O'Connor worked on terrain synthesis for an automatic radar map-matching guidance system. At Electronics Research Lahoratories he studied and worked on simulating low-frequency (0 to 0.1 Hz) noise of radar glint in the Nike missile system. He worked on the man-rating of the Titan III launcb vehicle at the Martin Co., and on the spacecraft attitude control propellent budget and structural oscillations (POGO) of the Apollo-Saturn launch vehicle for Bellcomm. In 1972 Mr. O'Connor joined Bell Lahoratories, where until 1973 he worked on the maintenance aspects of electromecbanical switching machines. From 1973 to 1976 he was involved with various traffic-related problems of toll switching machines. From 1976 until 1982 be worked on SONDS, and from 1982 until the present has been studying tbe project management aspects of the Dynamic Nonhierarchical Routing network.